AD-A225 698

ARI Research Note 90-34

DTIC FILE COPY

Creating Algorithms as an Aid to Judgment Part II

Sarah Lichtenstein and Anna Gilson Weathers

Perceptronics, Inc.

for

Contracting Officer's Representative
Michael Drillings

Basic Research
Michael Kaplan, Director

June 1990

FIFTY YEARS OF SERVICE 1940 - 1990

United States Army
Research Institute for the Behavioral and Social Sciences

Approved for public release; distribution is unlimited.

90

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Technical Director

JON W. BLADES
COL, IN
Commanding

Research accomplished under contract for the Department of the Army

Perceptronics, Inc.

Technical review by

Nehama Babin

X

A-1

NOTICES

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS --
2a. SECURITY CLASSIFICATION AUTHORITY --	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE --	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) PDRIR-1129-9-87	5. MONITORING ORGANIZATION REPORT NUMBER(S) ARI Research Note 90-34
6a. NAME OF PERFORMING ORGANIZATION Perceptronics, Inc. 6b. OFFICE SYMBOL (If applicable) --	7a. NAME OF MONITORING ORGANIZATION U.S. Army Research Institute
6c. ADDRESS (City, State, and ZIP Code) 6271 Variel Avenue Woodland Hills, CA 91367	7b. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600
8a. NAME OF FUNDING/SPONSORING ORGANIZATION U.S. Army Research Institute for the Behavioral and Social Sciences 8b. OFFICE SYMBOL (If applicable) BR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER MDA903-83-C-0198
8c. ADDRESS (City, State, and ZIP Code) 5001 Eisenhower Avenue Alexandria, VA 22333-5600	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO.: 61102B PROJECT NO.: 74F TASK NO.: WORK UNIT ACCESSION NO.:

11. TITLE (Include Security Classification)
Creating Algorithms as an Aid to Judgment: Part II

12. PERSONAL AUTHOR(S)
Lichtenstein, Sarah; and Weathers, Anna Gilson (Perceptronics, Inc.)

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Interim	FROM 83/04 TO 87/09	1990, June	18

16. SUPPLEMENTARY NOTATION
Contracting Officer's Representative, Michael Drillings

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Knowledge retrieval, Decomposed quantitative judgments

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

> Previous research has shown that subjects improve their performance in estimating unknown quantities when they are asked to write algorithms about these quantities. These algorithms are short combination rules whereby the subjects decompose the unknown quantities into a series of other quantities that they can more easily estimate and then use these estimates to arrive at an estimate of the requested quantity. This paper extends this research by exploring the effectiveness of asking subjects to work the algorithm they create not once, but twice, first using low estimates for the component parts and again using high estimates for the component parts. This study shows that subjects can be trained to create algorithms, but that the use of algorithms does not improve their performance to any great degree. When successful, the technique of creating an algorithm and using it to make two estimates led to very accurate performance (93% correct). But the technique also badly misled subjects on some occasions, bringing down the overall performance.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL Michael Drillings	22b. TELEPHONE (Include Area Code) (202) 274-8722 22c. OFFICE SYMBOL BR

DD Form 1473, JUN 86 Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

CREATING ALGORITHMS AS AN AID TO JUDGMENT: PART II

CONTENTS

LIST OF TABLES

CREATING ALGORITHMS AS AN AID TO JUDGMENT: PART II

INTRODUCTION

In previous research (Lichtenstein, MacGregor & Slovic, 1987) we have shown that subjects improve their performance in estimating unknown quantities when they are asked to write algorithms about these quantities. These algorithms are short combination rules whereby the subjects decompose the unknown quantities into a series of other quantities that they can more easily estimate and then recompose these estimates to arrive at an estimate of the requested quantity. For example, subjects may, using an algorithm, estimate the number of pounds of potato chips consumed yearly in the U.S. by estimating the number of pounds consumed by the average person per year and multiplying that estimate by the population of the U.S.

The present paper extends this research by exploring the effectiveness of asking subjects to work the algorithm they create not once, but twice, first using low estimates for the component parts and again using high estimates for the component parts.

METHOD

Subjects. The subjects were 276 people who answered an ad in the university student newspaper. They came to a large campus classroom at any time during the designated day to perform the experiment and were paid for their participation.

Stimuli. Nine questions were chosen for use in the experiment; these questions are shown in Table 1. Each question suggested to us a fairly simple algorithm that could be used to help in arriving at an estimate of the quantity requested.

Insert Table 1 about here

Table 1

Questions Used in the Experiment, with High and Low Answers.

	Correct Answer	High Answer	Low Answer	X^a
CivEmpl: How many U.S. Government civilian employees were there in August 1982?	2,908,025	5,816,050	1,454,013	2.
Chips: How many pounds of potato chips were consumed in the U.S. in 1982?	972,300,000	2,430,750,000	388,920,000	2.
Unempl: How many people received Oregon unemployment benefits in 1982?	188,000	564,000	63,000	3.
Taxes: In 1981 how much money did the state of Oregon collect in all kinds of taxes (including income tax, coporate income tax, gas tax, license fees, etc.)?	$1,608,423,000	$4,021,058,000	$643,369,000	2.
Horse: What is the world record speed for a thoroughbred racing horse to race 4 miles?	7 min 10 4/5 sec	10 min 46 1/5 sec	4 min 47 1/5 sec	1.
Car: How many minutes did it take for the 1982 winner of the Daytona 500 (a stock-car race) to complete the 500-mile race (rounded to nearest minute)?	195	390	98	2
SqMiles: How large is Oregon in square miles?	97,073	291,219	32,358	3
Popes: How many Roman Catholic popes have there been?	269	807	90	3
Counties: How many counties are there in the United States? (Alaska calls its counties "buroughs"; Louisiana calls its counties "parishes." Include these in your answer.)	3,077	6,154	1,539	2

Note: The abbreviated titles were not shown to the subjects.

[a]The factor used to arrive at the high and low answers.

Design and instructions. Each subject received all nine questions, each on a spearate page. For each question, an answer was provided. Half the time, this answer was much higher than the correct answer; half the time, it was too low. The subjects were assured that the answer provided was wrong; it was their task to decide whether the given answer was too high or too low.

The high and low answers used in the experiment are shown in Table 1, along with the correct answer and the factor used to arrive at the wrong answers. These factors were chosen, on the basis of pretests, so that approximately sixty percent of the students would be able to decide correctly whether our answer was too high or too low.

After making this assessment for the first three questions, the subjects received instructions in how to create an algorithm to aid them. They were instructed to build an estimate of the true answer to the question from facts they already knew or could estimate and to compare their estimate with our answer to see if our answer was too high or too low. The instructions gave three examples of algorithms. An easy algorithm was shown for the question, "How tall is the Empire State Building?", based on estimates of the number of floors and the height of each floor. Two more complex algorithms were then shown, both for the question, "What was the total attendance at all major league baseball games in 1983?" One of the algorithms was built from estimates of the number of teams, the number of games each team plays per year, and the average attendance per game. The other algorithm, for the same question, was based on the average yearly attendance per team and the number of teams.

After reading these instructions at their own pace, the subjects were given three more questions and asked to create an algorithm for each one before deciding whether our answer was too high or too low.

Before the final three questions, subjects were given a further set of instructions in which they were asked to construct two estimates for each question, one giving the lowest reasonable answer and one giving the highest reasonable answer. Each of the examples previously used was repeated here, showing the use of low and high estimates.

The instructions to subjects are shown in the Appendix.

Six forms of the questionnaire were prepared, such that, across the forms, each question appeared equally often with a high and a low answer and each question/answer combination appeared equally often in each third of the questionnaire. For each form, there was at least one but no more than two high answers in each third of the questionnaire and at least four but no more than five high answers overall.

RESULTS

Over all the data, there was no significant improvement in the correctness of the decisions made by subjects. In the first part, the subjects made 66.5% correct decisions, in the second part, 69.3%, and in the last part, 69.2%.

Each algorithm was coded by the experimenters according to whether the subject had successfully written an algorithm. The codes for the second part were:

N: No algorithm.

K: The subject did not write an algorithm, claiming to know the correct answer.

1: One algorithm.

For the last part, the codes were:

N: No algorithm.

K: The subject did not write an algorithm, claiming to know the correct answer.

1: One algorithm.

2: Two algorithms (i.e., one algorithm used twice, to produce two estimates)

For each case of two algorithms, the algorithms were further coded according to the relationship between the two estimates arrived at by the subject and the answer given:

2H: Both estimates higher than the given answer.

2L: Both estimates lower than the given answer.

2S: One estimate was higher and one estimate was lower than the given answer (S stands for "Split").

2U: Impossible for the coders to tell where the estimates stood in relation to the answer (U stands for "Unknown").

The results based on these codes is shown in Table 2. The results for the second part replicate our previous findings: Most of the time (85%) the subjects were able to write an algorithm, and their percentage correct was modestly higher when they could write an algorithm than when they could not.

Insert Table 2 about here

The results for the third part show that the subjects were usually (80%) able to create an algorithm and use it to arrive at two estimates.

Table 2

Results Based on Type of Algorithm

Second Part: One Algorithm		
Code	Frequency	Percent Correct
Missing Data	4	-
K	5	100
N	115	62
1	704	70

Third Part: Two Algorithms		
Code	Frequency	Percent Correct
Missing Data	11	-
K	3	100
N	60	60
1	96	63
2	659	71
2H given low	187	93
2H given high	45	4
2L given high	222	93
2L given low	78	19
2S	113	53
2U	14	57

When they did so, their decisions were more likely to be correct than when they did not make two estimates. However, a further analysis of the codes indicates a more complex situation. When both of the subject's estimates are higher than the given answer, it is natural that the subjects will decide that the given answer is too low. This is quite appropriate when the given answer is, in fact, too low. Subject's decisions were 93% accurate in this situation. But occasionally the subjects produced two estimates larger than the given answer even though the given answer was, in fact, too high. In this situation the subjects were seriously mislead by their estimates, getting only 4% of their decisions correct. The same reasoning applies in the reverse for two estimates that are lower than the given answer; the results are also parallel. When the subjects' two estimates are split, that is, when one is higher than the given answer and the other is lower, the entire effort gives no guidance to the subjects; in this situation the subjects were correct in only 53% of their decisions.

DISCUSSION

The present study shows that subjects can be trained to create algorithms, but that the use of algorithms does not improve their performance to any great degree. When successful, the technique of creating an algorithm and using it to make two estimates led to very accurate performance (93% correct). But the technique also badly mislead subjects on some occasions, bringing down the overall performance. Of course, subjects have no way of knowing, when they make two estimates, whether they are in the "successful" situation or the misleading situation. Thus, the method cannot be recommended unequivocably as a decision aid.

References

Lichtenstein, S., MacGregor, D., & Slovic, P. (1987). *Creating algorithms as an aid to judgment* (Report No. 87-5). Eugene, OR: Decision Research.

APPENDIX

Assessing Quantitative Facts

In this task we will present you with some questions, like "What is the world record time to run a mile?" For each question we provide an answer, like "8 minutes." EVERY ANSWER WE GIVE YOU IS WRONG. Your first task is to decide whether the answer provided is wrong because it is too LARGE a number, that is, too high, or because it is too SMALL a number, that is, too low.

The questions are straightforward; we have not used any "trick" items. We found the answers in various almanacs and the like, and then changed the answers, either by raising them or by lowering them. Some of the answers on your form of the questionnaire are too high; others are too low. In the first part there are only three questions. So please take your time on each one. Think hard about it before deciding whether the answer is too high or too low.

Your second task in this first part is to assess the probability that your decision is correct. Suppose you decided that our answer is too high. Then in this second task we want you to tell us the probability that the answer is, indeed, too high. If you decided that an answer is too low, now we want you to give us the probability that it is, indeed, too low.

This probability is stated in terms of "percent chances." It is a measure of the confidence you have in the correctness of your decision. If you are totally uncertain, so that you could just as well have decided with a flip of a coin, then you have a 50% chance of being right. If you are absolutely certain that you are right, as certain as you are of knowing your own name, then you are 100% certain. A response of 60%, for example, means that there are 60 chances out of 100, or 6 chances out of 10, that you made the right decision. Your answer in this second task should always be a number between 50% and 100%, inclusive. DO NOT USE A NUMBER SMALLER THAN 50 OR GREATER THAN 100.

You can start this task as soon as you are sure that you understand the instructions. Feel free to ask questions.

MORE INSTRUCTIONS

Next, you'll get three more questions, very like the three you just did. Again, we are giving you an answer that is WRONG. Again, we are asking for your judgment: Is the answer we give too high or too low? Again, after you make this decision, you should give us a number between 50 and 100 to express your confidence that your decision is correct.

But this time we want you to use a particular method for evaluating the answer. In a nutshell, you will use your own knowledge of related facts to construct your own answer, then compare your answer with ours.

The method will be clearer with a couple of examples. First, a very simple example:

How tall is the Empire State Building in New York City (excluding the TV antenna on top)?

Our answer: 2500 feet

Here's how to use the method: Forget our answer for a moment, and construct your own. What do you know that's relevant? Perhaps you can estimate the number of stories in the Empire State Building. Say, about 100. And you can give a reasonable estimate of the height of an average story, say about 10 feet. These two facts or estimates drawn from your own knowledge can be put together to get an estimate of the target quantity:

100 stories times 10 feet per story equals 1000 feet, height of building.

Your estimate, 1000 feet, is much lower than our answer. So our answer must be too high.

That's all there is to it. Search your memory and use your common sense to get facts or estimates that are relevant. Put these numbers together using simple arithmetic to arrive at your own estimate. Compare your estimate with our answer.

In fact, the Empire State Building is 1250 feet tall and has 102 stories. Thus in fact the average story is $1250 \div 102 = 12.25$ feet. Notice that your estimate of 10 feet per story was a bit off. Nevertheless, the method worked okay, because the answer we gave you was very much off. If you make small errors in your approximations you'll probably still do okay.

Now let's take an example that requires <u>several</u> facts or estimates.

What was the total attendance at all major league baseball games in 1983?

Our answer: 15,186,000

For our first try at this, we'll use four estimated facts:

1. There are about 30 major league teams.
2. Each team plays about 150 games per year.
3. Each game is played by 2 teams.
4. The average attendance at any one game is about 15,000.

Put these all together. Thirty teams times 150 games is 4500 team appearances per year. But since each game requires two teams, $4500 \div 2 =$ 2250 total games per year. Games per year times average attendance per game equals total attendance:

2250 times 15,000 = 33,750,000

Our new estimate is 33,750,000. That is much larger than the answer provided (15,186,000), so we conclude that the answer provided is too low.

(In fact, there are 26 major league teams and each team plays 162 games per year, and the average attendance per game is 21,632, but we were close enough in our estimates. The correct answer for the 1983 attendance is 45,557,582.)

There is usually more than one way to approach these questions. For example, suppose I don't have a good idea of how many teams there are, and I don't know how many games they play, but I remember reading that one team had a home attendance, for the year, of less than one million; the article implied that this was very low. So the average attendance for one team must be above one million. I guess it may be 1.5 million (1,500,000). It would have to be at least that high for that article I read to make such a big deal about falling below one million.

But how many teams are there? I remember there are two leagues. Each league must have at least 10 teams, for a total of 20 teams. I don't think they each have as many as 20 teams, for a total of 40 teams. Let's try an estimate a bit below the middle of that range, say 25 teams. Twenty five times 1.5 million attendance for each team gives a total attendance of 37,500,000. That is much higher than the provided answer. Even if I had used my low guess for the number of teams (20), I still would have gotten an estimate larger than the one provided. So it looks like the answer given is too small.

For each of the following three questions and answers, use the method explained above. Use simple arithmetic, some relevant facts or estimates from your own knowledge, and common sense to arrive at your own estimate of the answer. Then compare your answer with the one given.

Please write out enough words and numbers so that we can understand your approach. Try to make it legible and clear (but you don't need to write us a novel).

Take your time. We're giving you only three questions in this part so you can concentrate and do a careful job on each one.

Don't forget to give a confidence rating, too. The rating must be a number from 50 (for complete lack of confidence) to 100 (for utter certainty).

Now that you know this method, please do NOT go back to change any of your previous answers.

You can start as soon as you understand these instructions. Feel free to ask questions.

YET MORE INSTRUCTIONS

Finally, there are three more questions. For these, as for the last three, we would like you to construct estimates using facts you know. But this time we would like you to construct <u>two</u> estimates, one which gives the lowest reasonable answer and one which gives the highest reasonable answer.

Here's how you might use this method on the Empire State Building problem:

How tall is the Empire State Building?

Our Answer: 2500 feet.

Again, you try estimating the number of stories and the height of each story. But this time you do it twice.

First, make <u>low</u> estimates. You might say that you know there are at least 80 stories in the building. And it wouldn't make sense for the stories to be less than eight feet high, because there has to be room for the floor, the lights, and a 6-foot person. So the lowest the answer could be is:

80 stories times 8 feet per story = 640 feet high.

Second, use <u>high</u> estimates. There could be as many as 150 stories, and each one might be 15 feet high, but not more than that. So the tallest the Empire State Building could be is:

150 stories times 15 feet per story = 2250 feet high.

For this example, your low estimate is 640 feet and your high estimate is 2250 feet. Even your high estimate is smaller than our answer. So you conclude that our answer is too high.

Notice that you could have estimated that each story was 3 feet high or that there were 3000 stories. This would not have made sense. It won't help you to calculate the largest and smallest <u>possible</u> answers. Instead, use the largest and smallest <u>reasonable</u> answers.

You can evaluate the baseball example by the same methods. Recall the question and our answer:

What was the total attendance at all major league baseball games in 1983?

Our Answer: 15,186,000

For the lowest reasonable estimate, you might suppose that:

1) There are only 15 major league teams.
2) Each team plays 120 games per year.
3) Two teams play each game.
4) The average attendance at any one game is about 12,000.

So 15 teams play 120 games each for 1800 team-games. But two teams play each game, so there are 1800 ÷ 2 = 900 games per season. Finally, 900 games times 12,000 attending fans = 10,800,000 total attendance.

Then, for the highest reasonable estimate, you might suppose that:

1) There are 45 teams.
2) Each team plays 200 games per year.
3) Two teams play each game.
4) The average attendance at a game is 30,000.

Then you get:
45 times 200 = 9000
9000 ÷ 2 = 4500
4500 times 30,000 = 135,000,000 total attendance.

For this example, your low estimate (10,800,000) is lower than our answer, and your high estimate (135,000,000) is higher than our answer. This makes your decision more complicated. In such a case, review your estimates and revise them if you wish. Or just use your common sense to decide whether our answer is too high or too low.

If you know one component of your calculations, you do not need to vary it. For example, if you are sure that there are 26 teams and that each team plays 162 games per year, then your high and low estimates will vary only in the average attendance:

26 times 162 = 4212
4212 ÷ 2 = 2106 games per year
Games per year times average attendance = total attendance:

HIGH: 2106 times 30,000 = 63,180,000

LOW: 2106 times 12,000 = 25,272,000

Now both your high and your low estimates are larger than our answer, so you can reasonably conclude that our answer is too low.

As before, write out enough words and numbers so that we can understand your approach. In addition, please circle your two final estimates to help us find them when reading your paper. Also, continue to give a confidence rating, from 50 to 100.

Please do not return to any of your earlier answers now that you know this second method.

After these next three questions, there is one more short page. That completes the experiment. Return the materials and sign for your payment. The experimenter will check to see that you completed everything, that your handwriting is reasonably legible, and that all of your confidence ratings are between 50 and 100 (inclusive). When you finish, take a moment to check these things, too.

You can start as soon as you understand these instructions. Feel free to ask questions.

Thank you for your participation.